2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS) | 979-8-3503-1894-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/FOCS57990.2023.00029

When Does Adaptivity Help for Quantum State Learning?

Sitan Chen
SEAS
Harvard University
Cambridge, USA
sitan@seas.harvard.edu

Brice Huang
EECS
MIT
Cambridge, USA
bmhuang@mit.edu

Jerry Li
Machine Learning Foundations
Microsoft Research
Redmond, USA
jerrl@microsoft.com

Allen Liu
EECS
MIT
Cambridge, USA
cliu568@mit.edu

Mark Sellke
Department of Statistics
Harvard University
Cambridge, USA
msellke@fas.harvard.edu

***Abstract*—We consider the classic question of state tomography: given copies of an unknown quantum state $\rho \in \mathbb{C}^{d\times d}$, output $\widehat{\rho}$ which is close to ρ in some sense, e.g. trace distance or fidelity. When one is allowed to make coherent measurements entangled across all copies, $\Theta(d^2/\varepsilon^2)$ copies are necessary and sufficient to get trace distance ε [18], [29]. Unfortunately, the protocols achieving this rate incur large quantum memory overheads that preclude implementation on near-term devices. On the other hand, the best known protocol using incoherent (single-copy) measurements uses $O(d^3/\varepsilon^2)$ copies [24], and multiple papers have posed it as an open question to understand whether or not this rate is tight [6], [18]. In this work, we fully resolve this question, by showing that any protocol using incoherent measurements, even if they are chosen adaptively, requires $\Omega(d^3/\varepsilon^2)$ copies, matching the upper bound of [24]. We do so by a new proof technique which directly bounds the "tilt" of the posterior distribution after measurements, which yields a surprisingly short proof of our lower bound, and which we believe may be of independent interest.**

While this implies that adaptivity does not help for tomography with respect to trace distance, we show that it actually does help for tomography with respect to infidelity. We give an adaptive algorithm that outputs a state which is γ-close in infidelity to ρ using only $\widetilde{O}(d^3/\gamma)$ copies, which is optimal for incoherent measurements. In contrast, it is known [18] that any nonadaptive algorithm requires $\Omega(d^3/\gamma^2)$ copies. While it is folklore that in 2 dimensions, one can achieve a scaling of $O(1/\gamma)$, to the best of our knowledge, our algorithm is the first to achieve the optimal rate in all dimensions.

***Index Terms*—Quantum learning, quantum state tomography, adaptive algorithm, single-copy measurements**

I. Introduction

In this paper, we consider the problem of *quantum state tomography*. Here, the algorithm is given n copies of an unknown d-dimensional quantum mixed state ρ, and the goal is to output $\widehat{\rho}$ which is close to ρ in some sense.[1] There are several standard ways to measure closeness used in the literature. Arguably the two most important and well-studied notions are that of *trace distance* and *infidelity*[2], which are the natural quantum analogues of total variation distance and Hellinger distance, respectively. From a mathematical point of view, tomography is arguably the most fundamental quantum state learning task, and can also be seen as the natural non-commutative generalization of the classical problem of estimating a discrete distribution from samples. From a practical point of view, state tomography has many applications to the verification of quantum technologies [5].

In the *coherent* setting, that is, when the learner is allowed to make arbitrary measurements to the product state $\rho^{\otimes n}$, the situation is very well-understood. [18] demonstrated that for this problem, $n = O(d^2\log(d/\varepsilon)/\varepsilon^2)$ copies suffice to learn a state to error ε in trace distance, and $n = \Omega(d^2/\varepsilon^2)$ copies are necessary. Concurrently, [29] removed the logarithmic factor in the upper bound, thus establishing that the optimal rate for this problem is $n = \Theta(d^2/\varepsilon^2)$. The aforementioned paper of [18] also showed that, up to a single logarithmic factor, $\Omega(d^2/\gamma)$ copies are sufficient and necessary to learn a state to infidelity γ. Tight bounds are also known in many other settings, including for low rank states [18], [29], [30], as well as for a variety of related testing problems [3], [28].

However, in virtually all cases, the upper bounds that achieve the optimal rates require heavily entangled measurements. These measurements require all n copies of ρ to be prepared simultaneously, which—in addition to the complexity of preparing the measurements themselves—renders such approaches currently impractical. In light of this, there has been a recent surge of interest in the *incoherent* setting. Here, the algorithm is restricted to only making measurements of a single copy of ρ at a time. While such measurements are strictly weaker than general entangled measurements, they are much more practical and can be performed on real world quantum computers [21].

One of the main difficulties that arises when dealing with incoherent measurements is understanding the power of *adaptivity*. In general, an algorithm that uses incoherent measurements can sequentially measure each copy of ρ, and

SC was supported in part by NSF Award 2103300. BH was supported in part by an NSF Graduate Research Fellowship, a Siebel scholarship, NSF awards CCF-1940205 and DMS-1940092, and NSF-Simons collaboration grant DMS-2031883. AL was supported in part by an NSF Graduate Research Fellowship, a Fannie and John Hertz Foundation Fellowship and NSF awards CCF-1453261 and CCF-1565235. MS was supported in part by an NSF Graduate Research Fellowship, a Stanford Graduate Fellowship, and NSF grant CCF-2006489.

[1] Here and throughout the introduction, we will consider the setting where the success probability is some fixed constant.

[2] The fidelity between ρ, σ is $F(\rho,\sigma) = \mathsf{tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2$, and the infidelity is $1 - F(\rho,\sigma)$.

979-8-3503-1894-4/23/$31.00 ©2023 IEEE
DOI 10.1109/FOCS57990.2023.00029

moreover can choose how to measure the i-th copy of ρ based on the results of the previous $i-1$ measurements. We say such an algorithm is *adaptive*. In contrast, a *nonadaptive* algorithm must specify all of its measurements ahead of time.

Clearly, adaptive measurements are strictly more general than nonadaptive ones. On the other hand, to date, it has been unclear whether or not adaptivity actually yields better rates for any natural quantum learning problems. In large part, this is because until recently, it was not known how to obtain tight lower bounds against adaptive algorithms. Existing lower bound frameworks in the literature could only prove lower bounds against nonadaptive measurements; such algorithms are significantly easier to work with, and existing techniques from the classical learning literature could be adapted to the quantum setting.

However, starting with recent work of [6], there have been a flurry of works giving lower bounds against adaptive algorithms for a number of natural quantum learning problems. Now, lower bounds are known for a number of testing problems, such as mixedness testing and identity testing [6], [9], [10], shadow tomography [7], [8], [22], purity testing [1], [8]. In all of these cases, the lower bound against adaptive algorithms matches the best nonadaptive bound (up to polylogarithmic factors), so for all of these problems, adaptivity is not necessary to achieve optimal rates.

In contrast, for state tomography, progress has been slow in understanding the power of adaptivity. It is known that nonadaptively, $\Theta(d^3/\varepsilon^2)$ copies are necessary and sufficient for learning to error ε in trace distance [18], [24], and $\Theta(d^3/\gamma^2)$ copies are necessary and sufficient to learn to infidelity γ [18]. Similarly, as mentioned above, tight rates are known for the fully coherent setting (up to log factors). But nothing nontrivial is known in the adaptive setting: prior to our work, we did not know any adaptive algorithms that improve upon the best known nonadaptive rates, nor did we know any lower bounds against adaptive algorithms which were better than those known against coherent algorithms. Indeed, understanding the power of adaptive but incoherent measurements for quantum tomography has been posed as an open question by multiple papers [6], [18].

This problem is interesting on both fronts. From the lower bounds perspective, it seems that fundamentally new ideas are necessary to develop lower bounds for tomography. This is because all previous lower bound techniques for adaptive algorithms only held for testing problems, or via reductions to testing problems. These techniques will consequently fail for tomography, which is fundamentally a learning problem. From the upper bounds perspective, tomography in infidelity is one of the few natural candidates for a quantum learning problem where adaptive algorithms are provably better than nonadaptive ones. Indeed, it is folklore that in 2 dimensions, there are adaptive algorithms which achieve a scaling of $O(1/\gamma)$, although we could not find a reference to a full proof of this fact.

a) Our Results.: In this paper, we fully resolve the copy complexity of quantum state tomography with (potentially adaptively chosen) incoherent measurements, for both learning in trace distance and learning in infidelity. For trace distance, our main result is the following:

Theorem I.1 (Informal, see Theorem III.11)**.** *With incoherent measurements, $\Omega(d^3/\varepsilon^2)$ copies are necessary for tomography of d-dimensional states to ε error in trace distance.*

This matches the upper bound given by nonadaptive algorithms, and therefore for this problem, adaptivity provably does not help. In contrast, for learning in infidelity, our main result is the following:

Theorem I.2 (Informal, see Theorem VI.1)**.** *With adaptive, incoherent measurements, $\widetilde{O}(d^3/\gamma)$ copies[3] are sufficient for tomography of d-dimensional states to γ infidelity.*

Because trace distance and fidelity are related via $1 - F(\rho,\sigma) \leq \|\rho-\sigma\|_{\mathsf{tr}} \leq \sqrt{2(1-F(\rho,\sigma))}$ [16], Theorem I.1 implies that $\Omega(d^3/\gamma)$ copies are *necessary* for tomography in infidelity, so up to polylogarithmic factors, our upper bound in Theorem I.2 settles the complexity of tomography in infidelity with incoherent measurements. In contrast, [18] showed that $\Omega(d^3/\gamma^2)$ copies are necessary with nonadaptive measurements, so the rate in Theorem I.2 is provably better than what can be achieved with nonadaptive measurements, and thus provides arguably the first natural instance of a separation between the power of adaptive versus nonadaptive measurements.

b) Our Techniques.: Here, we give a very high level discussion of our proof techniques for Theorems I.1 and I.2. For a more detailed discussion, see Section II.

We begin with the lower bound. The main difficulty with proving lower bounds for state tomography with incoherent measurements is that essentially all the existing lower bound frameworks in the literature were fundamentally for testing problems. In such settings, it suffices to demonstrate hardness for a point-versus-mixture distinguishing task, where the goal is to distinguish between the case where the unknown mixed state is a single point versus the case where it is drawn from a mixture over alternate hypotheses. Such a setup is mathematically nice because the resulting likelihood ratios have a (relatively) simple multilinear form. However, for learning tasks, no such reduction exists; indeed, it is more natural to demonstrate hardness for a mixture-versus-mixture distinguishing task, but here the resulting likelihood ratios are much more complicated. Indeed, this phenomena seems to appear more generally in a variety of other (classical) learning settings, see e.g. [31].

We avoid this by directly bounding how much information the algorithm can learn from incoherent measurements, a technique which we believe may be of independent interest. We demonstrate that, for a carefully chosen prior on mixed states, the posterior distribution of the algorithm after $o(d^3/\varepsilon^2)$ incoherent measurements is anti-concentrated around the true

[3]We use $f = \widetilde{O}(g)$ to denote that $f = C \cdot g \cdot \log^c(g))$ for some absolute constants $c, C > 0$.

mixed state. It is perhaps surprising that we are able to directly bound the behavior of the posterior distribution, but it turns out that the "tilt" caused by the measurements can be upper bounded "by hand", and then the required anti-concentration follows from classical results in random matrix theory.

We now turn our attention to the upper bound. First, let us briefly discuss why adaptivity may help for learning in infidelity. At a very high level, learning in infidelity seems to correspond to learning the eigenvalues and eigenvectors of the density matrix to some degree of relative accuracy. However, nonadaptive algorithms are unable to do this easily: the small eigenspaces are hidden by the "noise" caused by the large eigenvalues. On the other hand, adaptive algorithms can learn the large eigenspaces, then project them away to reveal the information about the smaller eigenvalues.

The main challenge with this approach is dealing with the error accumulation in the infidelity as we iterate this process. In particular, we cannot exactly learn the large eigenspaces, and so we must control the error the projections incur. In constant dimensions, one can directly brute force the calculations, but in high dimensions, the calculations become intractable. Our main technical contribution for this part of the paper is a new technique to bound the fidelity between two noncommuting states, by carefully guessing a matrix square root for their symmetrized product.

c) Comparison to Prior Work.: As discussed above, our techniques are very different from the ones used to prove lower bounds for quantum *testing* in the incoherent measurements setting [1], [6]–[10], [12], [22]. Here we briefly mention the sole commonalities between our proof and these works.

First, we use the helpful abstraction of adaptive algorithms as given by decision trees. This is common to all aforementioned works on incoherent measurements, but here we clarify a common misconception: the tree representation is not so much a central technical ingredient as it is a notational necessity for keeping track of the internal state of the algorithm.

Secondly, similar to [9], in place of an ensemble based on Haar-random unitaries, we work with a suitable Gaussian approximation. While the former is the construction of choice for many known quantum property testing lower bounds, it was observed in [9] that the latter is more amenable to certain high-degree moment calculations that arise in the proofs of these lower bounds. We emphasize however that such moment calculations are not relevant to the current work, and the Gaussian ensemble turns out to be convenient for fairly different reasons, namely certain density and isotropy properties (see Section II-B).

The similarities between our techniques and previous work essentially end here. In particular, our main lower bound strategy (which allows us to directly reason about the change in the posterior distribution) is to our knowledge completely novel, and we believe it may be of independent interest.

d) Additional Related Work.: Apart from the aforementioned bounds for state tomography and quantum state testing, there have also been lower bounds in the incoherent setting when the measurements are partially adaptive or come from a set of bounded size [25], and when the measurements are Pauli [15].

We also note there is a large literature on understanding the power that adaptivity affords for tomography in infidelity in the *asymptotic* setting [4], [13], [20], [26]. They obtain rates that are linear in $1/\gamma$, but unlike our Theorem I.2, these results get some unspecified dependence on d and/or only apply to the regime of $d = O(1)$.

[32] gives upper and lower bounds for *pure state* tomography in a different setting where instead of getting copies of the unknown state, one has access to a unitary which prepares the state.

Finally, the recent work of [11] gives a constant-factor separation between non-adaptive and adaptive algorithms for quantum hypothesis selection, as well as a polynomial separation for a problem in which one is promised that an unknown state can be diagonalized in one of m known bases and would like to approximate the state in trace distance by measuring copies. Their separation is with respect to the parameters d and m, rather than ϵ as in our work.

e) Concurrent work.: Our proof of Theorem I.2 directly generalizes to give that $\widetilde{O}(dr^2/\gamma)$ measurements are sufficient when the unknown state has rank-r. In concurrent and independent work, Flammia and O'Donnell [14] also obtain, up to polylogarithmic factors, an upper bound of $O(dr^2/\gamma)$ for tomography with adaptive single-copy measurements. Their guarantee is somewhat stronger as their error is measured in quantum relative entropy, which is an upper bound on infidelity. They also obtain a bound of $\widetilde{O}(d^{3/2}r^{3/2}/\gamma)$ for the same problem where γ is given by Bures chi-squared divergence, which also upper bounds infidelity.

II. Technical Overview

a) Notation.: In addition to standard big-O notation, we sometimes also use the notation $f \lesssim g$ (resp. $f \gtrsim g$) to denote that there exists an absolute constant $C > 0$ such that $f \leq Cg$ (resp. $f \geq Cg$).

A. Preliminaries

Throughout, let ρ denote a quantum state. We recall that a quantum state is a psd Hermitian matrix in $\mathbb{C}^{d\times d}$ with trace 1.

a) Measurements.: We now define the standard measurement formalism, which is the way algorithms are allowed to interact with a quantum state ρ.

Definition II.1 (Positive operator valued measurement (POVM), see e.g. [27])**.** *A positive operator valued measurement $\mathcal{M}$ is a finite collection of psd matrices $\mathcal{M} = \{M_z\}_{z\in\mathcal{Z}}$ satisfying $\sum_z M_z = I_d$. When a state ρ is measured using $\mathcal{M}$, we get a draw from a classical distribution over $\mathcal{Z}$, where we observe z with probability $\mathrm{tr}(\rho M_z)$. Afterwards, the quantum state is destroyed.*

In this work we assume that all POVMs used are rank-1. It is a standard fact that this is without loss of generality (see e.g. [8, Lemma 4.8]). We will sometimes represent a sequence of

n measurement outcomes by $\boldsymbol{x} = (x_1, \ldots, x_n)$ to denote that in the i-th step, the outcome that was observed corresponds to a POVM element which is a scalar multiple of $x_i x_i^\dagger$.

b) Incoherent Measurements.: An algorithm using incoherent measurements operates as follows: given n copies of ρ, it iteratively measures the i-th copy using a POVM (which could depend on the results of previous measurements), records the outcome, and then repeats this process on the $(i+1)$-th copy. After having performed all n measurements, it must output an estimate of ρ based on the (classical) sequence of outcomes it has received. Formally, such an algorithm can be represented as a tree, see Definition III.1.

B. Lower Bound

We begin with a heuristic calculation that explains how the threshold d^3/ε^2 enters our lower bound. We assume heuristically that we have a discrete set $\mathcal{S}$ of quantum states $\rho \in \mathbb{R}^{d\times d}$ satisfying the following properties.

a) *Near-mixedness*: for all $\rho \in \mathcal{S}$, all eigenvalues of ρ lie in $[0.9/d, 1.1/d]$.
b) *Packing*: for all $\rho, \rho' \in \mathcal{S}$, $\|\rho - \rho'\|_{\mathsf{tr}} \geq 2\varepsilon$.
c) *Density*: for a $1 - o(1)$ fraction of $\rho_0 \in \mathcal{S}$, the neighborhood

$$N(\rho_0) = \{\rho \in \mathcal{S} : \|\rho - \rho_0\|_{\mathsf{tr}} \leq 100\varepsilon\}$$

has cardinality $|N(\rho_0)| \geq \exp(\Omega(d^2))$.
d) *Isotropy*: for a $1-o(1)$ fraction of $\rho_0 \in \mathcal{S}$, the distribution $\gamma(\rho_0) = \mathrm{unif}(N(\rho_0))$ is isotropic around ρ_0, in the sense that

$$\mathbb{E}_{\rho\sim\gamma(\rho_0)}[\rho] = \rho_0, \tag{1}$$

$$\mathbb{E}_{\rho\sim\gamma(\rho_0)}[(v^\dagger(\rho-\rho_0)v)^2] \lesssim \varepsilon^2/d^3, \quad \forall \|v\|_2 = 1. \tag{2}$$

(The below argument still works if (1) holds only approximately, but having it hold with equality simplifies the discussion.)

We expect this can be achieved by a random construction. The scale $\exp(\Omega(d^2))$ arises from the dimension of the ambient space of quantum states, and the right-hand side ε^2/d^3 of (2) is the scale we get if we replace $\gamma(\rho_0)$ with an isotropic distribution around ρ_0 with appropriate trace distance, for example that of $\rho = \rho_0 + \frac{100\varepsilon}{d} U^\dagger Z U$ for Haar random $U \in \mathbb{R}^{d\times d}$ and $Z = \mathrm{diag}(1, \ldots, -1, \ldots)$. We expect the density and isotropy conditions to hold for all $\rho_0 \in \mathcal{S}$ not too close to the boundary of the point cloud $\mathcal{S}$.

Assuming such $\mathcal{S}$ exists, we will show that it is not possible to learn ρ_0 sampled from the hard distribution $\mu = \mathrm{unif}(\mathcal{S})$. Because $\mathcal{S}$ is a 2ε-packing in trace distance, learning $\rho_0 \in \mathcal{S}$ to ε in trace distance amounts to recovering ρ_0.

The central idea of the proof is that, if ρ_0 satisfies the mixedness and isotropy conditions (a), (d), each measurement improves the log likelihood ratio of ρ_0 to $N(\rho_0)$ by at most $O(\varepsilon^2/d)$. So, if ρ_0 satisfies the density condition (c), after observing $n \ll d^3/\varepsilon^2$ measurements the improvement $\exp(O(n\varepsilon^2/d))$ in the likelihood ratio is not enough to overcome the cardinality $\exp(\Omega(d^2))$ of $N(\rho_0)$. Thus the posterior distribution of ρ_0 still places much more mass on $N(\rho_0)$ than ρ_0 itself.

More concretely, let the observations (from adaptive POVMs) of ρ_0 be $\boldsymbol{x} = (x_1, \ldots, x_n)$ and denote the posterior distribution of ρ_0 by $\nu_{\boldsymbol{x}}(\cdot)$. Then,

$$\frac{\nu_{\boldsymbol{x}}(N(\rho_0))}{\nu_{\boldsymbol{x}}(\rho_0)} = \sum_{\rho\in N(\rho_0)} \prod_{i=1}^n \frac{x_i^\dagger \rho x_i}{x_i^\dagger \rho_0 x_i}$$

$$= |N(\rho_0)|\, \mathbb{E}_{\rho\sim\gamma(\rho_0)} \prod_{i=1}^n \frac{x_i^\dagger \rho x_i}{x_i^\dagger \rho_0 x_i} \tag{3}$$

$$\geq \exp\left\{\Omega(d^2) + \sum_{i=1}^n \mathbb{E}_{\rho\sim\gamma(\rho_0)} \log\left(1 + \frac{x_i^\dagger(\rho-\rho_0)x_i}{x_i^\dagger\rho_0 x_i}\right)\right\}, \tag{4}$$

where the last step uses Jensen's inequality. For suitably small ε, Taylor expanding $\log(1+x)$ gives

$$\mathbb{E}_{\rho\sim\gamma(\rho_0)} \log\left(1 + \frac{x_i^\dagger(\rho-\rho_0)x_i}{x_i^\dagger\rho_0 x_i}\right)$$

$$\geq \mathbb{E}_{\rho\sim\gamma(\rho_0)} \frac{x_i^\dagger(\rho-\rho_0)x_i}{x_i^\dagger\rho_0 x_i} - \mathbb{E}_{\rho\sim\gamma(\rho_0)} \frac{2}{3}\left(\frac{x_i^\dagger(\rho-\rho_0)x_i}{x_i^\dagger\rho_0 x_i}\right)^2$$

$$\geq -d^2\, \mathbb{E}_{\rho\sim\gamma(\rho_0)}[(x_i^\dagger(\rho-\rho_0)x_i)^2] \qquad \text{(by (a))}$$

$$\gtrsim -\varepsilon^2/d \qquad \text{(by (2)).}$$

The hypothesis $n \ll d^3/\varepsilon^2$ implies $n\varepsilon^2/d \ll d^2$, so (4) is $\exp(\Omega(d^2))$. Thus the posterior mass on ρ_0 is

$$\nu_{\boldsymbol{x}}(\rho_0) \leq \frac{\nu_{\boldsymbol{x}}(\rho_0)}{\nu_{\boldsymbol{x}}(N(\rho_0))} \leq \exp(-\Omega(d^2))$$

and learning ρ_0 is impossible.

However, it is hard to make the above approach rigorous, due to the difficulty of constructing a point set $\mathcal{S}$ with the required density and isotropy properties. We sidestep this issue by instead working with a continuous prior μ, namely a perturbation of the maximally mixed state $\frac{1}{d}I_d$ by a suitable trace-centered Gaussian matrix. In this approach, in lieu of showing $\nu_{\boldsymbol{x}}(N(\rho_0))/\nu_{\boldsymbol{x}}(\rho_0)$ is large, our goal will be to show $\nu_{\boldsymbol{x}}(B(\rho_0, C\varepsilon))/\nu_{\boldsymbol{x}}(B(\rho_0,\varepsilon))$ is large, where $B(\rho_0,\varepsilon)$ denotes a trace distance ball of radius ε and C is a large constant. In the new proof, $\gamma(\rho_0)$ will be μ restricted to $B(\rho, C\varepsilon)$, then **sub-sampled** to be transparently isotropic. Being able to sub-sample a measure in this way is a key advantage of the continuous setup.

The central idea of the proof is still that each measurement improves the log likelihood ratio by $O(\varepsilon^2/d)$, which is not enough to overcome the volume ratio $|B(\rho_0, C\varepsilon)|/|B(\rho_0,\varepsilon)| = C^{\Theta(d^2)}$, which serves as the continuous analogue of the cardinality of $N(\rho_0)$ from the earlier argument. Although sub-sampling incurs factors that decrease our estimate of $\nu_{\boldsymbol{x}}(B(\rho_0, C\varepsilon))/\nu_{\boldsymbol{x}}(B(\rho_0,\varepsilon))$ by a $\exp(\Omega(d^2))$ factor, this is overcome by setting C to be a suitably large constant.

C. Upper Bound

The first observation about infidelity compared to trace distance is that infidelity is much more sensitive to errors on smaller eigenvalues. For instance, consider two states $\rho = \mathrm{diag}(\rho_1, \ldots, \rho_d)$ and $\rho' = \mathrm{diag}(\rho'_1, \ldots, \rho'_d)$ where $\rho'_i = \rho_i + \Delta_i$ for all $i \in [d]$. When the Δ_i are sufficiently small (to ensure the following power series expansion is a good approximation), we have

$$1 - F(\rho, \rho') = 1 - \sum_{i=1}^{d} \sqrt{\rho_i(\rho_i + \Delta_i)}$$
$$\approx 1 - \sum_{i=1}^{d} \left(\rho_i - \frac{\Delta_i}{2} - \frac{\Delta_i^2}{4\rho_i}\right) = \sum_{i=1}^{d} \frac{\Delta_i^2}{4\rho_i}\,.$$

The ρ_i in the denominator means that it is necessary to get better accuracy on smaller eigenvalues. Heuristically, this means to get infidelity γ we should need to learn eigenvalues of size $\sim \sigma$ to accuracy $\sqrt{\gamma\sigma/d}$.

In the classical setting of estimating a discrete distribution, if we simply take the empirical point estimates, we naturally get finer additive accuracy for the lower probability points. However, in the quantum setting this is not the case because we do not know the right basis to measure in. If we run a nonadaptive tomography algorithm on an unknown state ρ with d^3/γ samples, we can (see Theorem V.4) learn a state $\hat{\rho}$ such that $\|\rho - \hat{\rho}\|_{\mathsf{op}} \leq \frac{\sqrt{\gamma}}{d}$. This would suffice for learning to infidelity γ if the eigenvalues of ρ are all $\Omega(1/d)$. However, if ρ has small eigenvalues, they can be dominated by the estimation error in $\hat{\rho}$.

Our learning algorithm overcomes this issue by running logarithmically many rounds. In each round, we run a standard tomography algorithm and argue that this accurately learns the large eigenvalues and corresponding eigenspace. Then we can essentially project out this eigenspace and restrict to the orthogonal complement in all future measurements using appropriately designed POVMs (see Definition V.6). This lets us focus on the smaller eigenvalues and learn those to finer accuracy. We can then repeat by running a tomography algorithm on the orthogonal complement, learning and projecting out the largest remaining eigenvalues and so on. At the end of the algorithm we will have learned projection matrices $\Pi_1 = B_1 B_1^\top, \ldots, \Pi_t = B_t B_t^\top$ (where B_i has columns forming an orthonormal basis of the corresponding subspace) where $t \sim \log(d/\varepsilon)$ and an estimator $\hat{\rho}$ that is diagonal in the basis given by $\{B_1, \ldots, B_t\}$. Let d_i be the dimension of B_i so $d = d_1 + \cdots + d_t$. In this basis, we can write

$$\hat{\rho} = \begin{pmatrix} \widehat{\rho_1} & & & \\ & \widehat{\rho_2} & & \\ & & \ddots & \\ & & & \widehat{\rho_t} \end{pmatrix}$$

$$\rho = \begin{pmatrix} \rho_1 & E_1 & & \\ E_1^\top & \rho_2 & E_2 & \\ & E_2^\top & \ddots & \\ & & & \rho_t \end{pmatrix}$$

where $\rho_i, \widehat{\rho_i} \in \mathbb{R}^{d_i \times d_i}$, $E_i \in \mathbb{R}^{d_i \times (d_{i+1} + \cdots + d_t)}$. The guarantees of our tomography algorithm give us appropriate scale-sensitive bounds on $\|\widehat{\rho_i} - \rho_i\|_{\mathsf{op}}$ and $\|E_i\|_{\mathsf{op}}$. We can then bound the infidelity between $\rho, \hat{\rho}$ in two stages. We first bound the infidelity coming from the differences on the diagonal blocks i.e. $1 - F(\hat{\rho}, \rho_{\mathrm{diag}})$ where $\rho_{\mathrm{diag}} = \mathrm{diag}(\rho_1, \ldots, \rho_t)$ (see Lemma VI.5). We then bound the infidelity coming from the off-diagonal blocks (see Lemma VI.6) This is the main technical component of the proof. The matrix square root that appears in the expression for fidelity can become very complicated and unwieldy. Instead, we bound it by carefully guessing a matrix that lower bounds the square root (see Lemma VIII.4).

III. Proof of Lower Bound

In this section we prove Theorem III.11, our hardness result for tomography in trace distance.

A. Lower Bound Framework

We begin by reviewing a standard framework for representing an adaptive algorithm as a tree.

Definition III.1 (Tree representation, see e.g. [8])**.** *Fix an unknown d-dimensional mixed state ρ. An algorithm for state tomography that only uses n incoherent, possibly adaptive, measurements of ρ can be expressed as a pair $(\mathcal{T}, \mathcal{A})$, where $\mathcal{T}$ is a rooted tree $\mathcal{T}$ of depth n satisfying the following properties:*

- *Each node is labeled by a string of vectors $\boldsymbol{x} = (x_1, \ldots, x_t)$, where each x_i corresponds to measurement outcome observed in the i-th step.*
- *Each node $\boldsymbol{x}$ is associated with a probability $p^\rho(\boldsymbol{x})$ corresponding to the probability of observing $\boldsymbol{x}$ over the course of the algorithm. The probability for the root is 1.*
- *At each non-leaf node, we measure ρ using a rank-1 POVM $\{\omega_x d \cdot xx^\dagger\}_x$ to obtain classical outcome $x \in \mathbb{S}^{d-1}$. The children of $\boldsymbol{x}$ consist of all strings $\boldsymbol{x}' = (x_1, \ldots, x_t, x)$ for which x is a possible POVM outcome.*
- *If $\boldsymbol{x}' = (x_1, \ldots, x_t, x)$ is a child of $\boldsymbol{x}$, then*

$$p^\rho(\boldsymbol{x}') = p^\rho(\boldsymbol{x}) \cdot \omega_x d \cdot x^\dagger \rho x\,. \tag{5}$$

- *Every root-to-leaf path is length-n. Note that $\mathcal{T}$ and ρ induce a distribution over the leaves of $\mathcal{T}$.*

$\mathcal{A}$ is a randomized algorithm that takes as input any leaf $\boldsymbol{x}$ of $\mathcal{T}$ and outputs a state $\mathcal{A}(\boldsymbol{x})$. The output of $(\mathcal{T}, \mathcal{A})$ upon measuring n copies of a state ρ is the random variable $\mathcal{A}(\boldsymbol{x})$, where $\boldsymbol{x}$ is sampled from the aforementioned distribution over the leaves of $\mathcal{T}$.

We also recall the definition of the Gaussian Orthogonal Ensemble (GOE) and define a trace-centered variant, which will be the basis of our hard distribution.

Definition III.2 (GOE, Trace-centered GOE)**.** *A sample $G \sim \mathrm{GOE}(d)$ is a symmetric matrix with independent Gaussians on and above the diagonal, with $G_{i,i} \sim \mathcal{N}(0, 2/d)$ and $G_{i,j} \sim$*

$\mathcal{N}(0, 1/d)$ *for* $i < j$. *A sample* $G' \sim \mathrm{GOE}^*(d)$ *is sampled by* $G' = G - \mathrm{tr}(G)$ *where* $G \sim \mathrm{GOE}(d)$.

The trace-centered GOE also features in the hard instance of [9] for state certification, but as discussed in Section I, our reason for working with this ensemble is very different from that of [9]. We recall the following standard fact about extremal eigenvalues of the GOE matrix.

Lemma III.3 ([8, Lemma 6.2])**.** *If* $G \sim \mathrm{GOE}^*(d)$, *then* $\|G\|_{\mathsf{op}} \leq 3$ *with probability* $1 - e^{-\Omega(d)}$.

B. Construction of Hard Distribution

We construct the following hard distribution μ of quantum states. Let $U \subseteq \mathbb{R}^{d\times d}$ be the affine subspace of symmetric matrices with trace 1 and $U_0 \subseteq \mathbb{R}^{d\times d}$ be the linear subspace of symmetric matrices with trace 0. These spaces inherit the inner product of $\mathbb{R}^{d\times d}$, which defines Lebesgue measures Leb_U and Leb_{U_0} on them. Let $\sigma = \frac{1}{100}$ be a small constant. A sample $\rho \sim \mu$ is generated by

$$\rho = \frac{1}{d}(I_d + \sigma G),$$

where G is a sample from $\mathrm{GOE}^*(d)$ conditioned on $\|G\|_{\mathsf{op}} \leq 4$. Note that such matrices are clearly psd and thus valid quantum states. Concretely, μ has density (with respect to Leb_U)

$$\mu(\rho) = \frac{1}{Z}\exp\left(-\frac{d^3}{4\sigma^2}\|\rho - \frac{1}{d}I_d\|_F^2\right)\mathbb{1}\{\rho \in S_{\text{supp}}\}$$

$$S_{\text{supp}} = \left\{\rho \in U : \|\rho - \frac{1}{d}I_d\|_{\mathsf{op}} \leq \frac{4\sigma}{d}\right\}.$$

where Z is a normalizing constant. Further define a set of "good" states

$$S_{\mathsf{good}} = \left\{\rho \in U : \|\rho - \frac{1}{d}I_d\|_{\mathsf{op}} \leq \frac{3\sigma}{d}\right\},$$

which corresponds to the event $\|G\|_{\mathsf{op}} \leq 3$. Due to Lemma III.3, $\mu(S_{\mathsf{good}}) \geq 1 - e^{-\Omega(d)}$.

In the below proof, we will show that all $\rho_0 \in S_{\mathsf{good}}$ are hard to learn. The important property of S_{good} is that it is far from the boundary of $\mathrm{supp}(\mu) = S_{\text{supp}}$; this ensures that we can choose a suitable sub-sampling of μ in a neighborhood of ρ_0, which is isotropic around ρ_0.

Finally we record the following straightforward fact.

Lemma III.4. *For all* $\rho, \rho' \in S_{supp}$, $\exp(-4d^2) \leq \mu(\rho)/\mu(\rho') \leq \exp(4d^2)$.

Proof. For all $\rho \in S_{\text{supp}}$,

$$0 \leq \frac{d^3}{4\sigma^2}\|\rho - \frac{1}{d}I\|_F^2 \leq \frac{d^4}{4\sigma^2}\|\rho - \frac{1}{d}I\|_{\mathsf{op}}^2 \leq 4d^2. \qquad \square$$

C. Anticoncentration of Posterior Distribution

Fix a tomography algorithm $(\mathcal{T}, \mathcal{A})$ as in Definition III.1, and let $\mathcal{T}_\rho$ denote the distribution over observation sequences $\boldsymbol{x} = (x_1, \ldots, x_n)$ when $\mathcal{T}$ is run on state ρ. Note that for any states ρ, ρ' in the support of μ, the likelihood ratio

$$\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho'}}(\boldsymbol{x}) = \prod_{i=1}^n \frac{x_i^\dagger \rho x_i}{x_i^\dagger \rho' x_i}$$

is well defined, since μ is supported on full-rank matrices. Let $\nu_{\boldsymbol{x}}$ denote the posterior distribution of ρ given observations $\boldsymbol{x}$. The density ratio of any $\rho, \rho' \in S_{\text{supp}}$ under $\nu_{\boldsymbol{x}}$ is given by Bayes' rule, and equals

$$\frac{\nu_{\boldsymbol{x}}(\rho)}{\nu_{\boldsymbol{x}}(\rho')} = \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho'}}(\boldsymbol{x}) \cdot \frac{\mu(\rho)}{\mu(\rho')}.$$

So, for an arbitrary reference state $\rho' \in S_{\text{supp}}$ (below we take $\rho' = \rho_0$, the unknown true state) the density of $\nu_{\boldsymbol{x}}$ is

$$\nu_{\boldsymbol{x}}(\rho) = \frac{1}{Z_{\boldsymbol{x}}}\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho'}}(\boldsymbol{x})\mu(\rho)$$

$$Z_{\boldsymbol{x}} = \int_U \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho'}}(\boldsymbol{x})\mu(\rho)\ \mathrm{d}\mathsf{Leb}_U(\rho).$$

The main technical component of the proof is the following anti-concentration result for $\nu_{\boldsymbol{x}}$.

Definition III.5. *For* $\rho \in U$, *let* $B(\rho, \varepsilon)$ *denote the ball* $\{\rho' \in U : \|\rho' - \rho\|_{\mathsf{tr}} \leq \varepsilon\}$. *Similarly for* $\rho \in U_0$, *let* $B(\rho, \varepsilon) = \{\rho' \in U_0 : \|\rho' - \rho\|_{\mathsf{tr}} \leq \varepsilon\}$.

Theorem III.6. *Suppose* $d \gg 1$, $\varepsilon \leq \varepsilon_0$ *for an absolute constant* ε_0, *and* $n \ll d^3/\varepsilon^2$. *If* $\rho \in S_{\mathsf{good}}$ *and* $\boldsymbol{x} \sim \mathcal{T}_{\rho_0}$, *there is an event* $S_{\rho_0} \in \sigma(\boldsymbol{x})$, *with* $\mathbb{P}(\boldsymbol{x} \in S_{\rho_0}) \geq 1 - \exp(-d^2)$, *on which* $\nu_{\boldsymbol{x}}(B(\rho_0, \varepsilon)) \ll 1$.

Let C be a large constant we will set later and $\varepsilon_0 = \sigma/C$. The starting point of the proof of Theorem III.6 is the estimate

$$\frac{\nu_{\boldsymbol{x}}(B(\rho_0, C\varepsilon))}{\nu_{\boldsymbol{x}}(B(\rho_0, \varepsilon))} = \frac{\int_{B(\rho_0, C\varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x})\mu(\rho)\ \mathrm{d}\mathsf{Leb}_U(\rho)}{\int_{B(\rho_0, \varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x})\mu(\rho)\ \mathrm{d}\mathsf{Leb}_U(\rho)} \tag{6}$$

$$\geq \exp(-4d^2)\frac{\int_{B(\rho_0, C\varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x})\mathbb{1}\{\rho \in S_{\text{supp}}\}\ \mathrm{d}\mathsf{Leb}_U(\rho)}{\int_{B(\rho_0, \varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x})\mathbb{1}\{\rho \in S_{\text{supp}}\}\ \mathrm{d}\mathsf{Leb}_U(\rho)}, \tag{7}$$

where the second line uses Lemma III.4. Applying Lemma III.4 in this way amounts to replacing μ in the numerator of (6) with a measure that sub-samples it, and in the denominator with a measure that upper bounds it. Define the volumes

$$V_1 = \int_{B(0,1)} \mathrm{d}\mathsf{Leb}_{U_0}(\rho)$$

$$V_2 = \int_{B(0,1)} \mathbb{1}\{\|\rho\|_{\mathsf{op}} \leq 1/d\}\ \mathrm{d}\mathsf{Leb}_{U_0}(\rho).$$

We now separately bound the numerator and denominator of (7) in terms of these volumes, beginning with the denominator.

Lemma III.7. *If $\rho_0 \in S_{\mathsf{good}}$, there is an event $S_{\rho_0} \in \sigma(\boldsymbol{x})$, with $\mathbb{P}(\boldsymbol{x} \in S_{\rho_0}) \geq 1 - \exp(-d^2)$, on which*

$$\int_{B(\rho_0,\varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \mathbb{1}\{\rho \in S_{supp}\} \, \mathrm{dLeb}_U(\rho) \leq \exp(d^2)\varepsilon^{(d+2)(d-1)/2} V_1 \, .$$

Proof. Note that $\mathbb{E}_{\boldsymbol{x}\sim\mathcal{T}_{\rho_0}} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) = 1$. So

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{x}\sim\mathcal{T}_{\rho_0}} \int_{B(\rho_0,\varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \mathbb{1}\{\rho \in S_{\mathrm{supp}}\} \, \mathrm{dLeb}_U(\rho) \\
&= \int_{B(\rho_0,\varepsilon)} \mathbb{1}\{\rho \in S_{\mathrm{supp}}\} \, \mathrm{dLeb}_U(\rho) \\
&\leq \int_{B(\rho_0,\varepsilon)} \mathrm{dLeb}_U(\rho) = \varepsilon^{(d+2)(d-1)/2} V_1.
\end{aligned}$$

The exponent $(d+2)(d-1)/2$ comes from the fact that the space of symmetric matrices has dimension $d(d+1)/2$, so U has dimension $d(d+1)/2 - 1 = (d+2)(d-1)/2$. The result follows from Markov's inequality. □

Before bounding the numerator of (7), we define the set

$$N_*(\rho_0) = \left\{ \rho \in U : \|\rho - \rho_0\|_{\mathsf{op}} \leq \frac{C\varepsilon}{d}, \|\rho - \rho_0\|_{\mathsf{tr}} \leq C\varepsilon \right\},$$

which is an isotropic neighborhood of ρ_0. Let $\gamma(\rho_0)$ denote the uniform distribution on $N_*(\rho_0)$ (w.r.t. Leb_U). That is, for bounded measurable test function $f : U \to \mathbb{R}$,

$$\mathbb{E}_{\rho\sim\gamma(\rho_0)} f(\rho) = \frac{\int_U f(\rho)\mathbb{1}\{\rho \in N_*(\rho_0)\}\mathrm{dLeb}_U(\rho)}{\int_U \mathbb{1}\{\rho \in N_*(\rho_0)\}\mathrm{dLeb}_U(\rho)}.$$

Lemma III.8. *If $\rho_0 \in S_{\mathsf{good}}$ and $\varepsilon \leq \varepsilon_0$, then*

$$\begin{aligned}
&\int_{B(\rho_0,C\varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \mathbb{1}\{\rho \in S_{supp}\} \, \mathrm{dLeb}_U(\rho) \\
&\geq (C\varepsilon)^{(d+2)(d-1)/2} \, \mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \right] V_2.
\end{aligned} \tag{8}$$

Proof. Because $\rho_0 \in S_{\mathsf{good}}$, we have $\|\rho_0 - \frac{1}{d} I_d\|_{\mathsf{op}} \leq \frac{3\sigma}{d}$. Because $\varepsilon \leq \varepsilon_0 = \sigma/C$, if ρ satisfies $\|\rho - \rho_0\|_{\mathsf{op}} \leq \frac{C\varepsilon}{d}$ we have the implication chain

$$\begin{aligned}
&\|\rho - \rho_0\|_{\mathsf{op}} \leq \frac{C\varepsilon}{d} \Rightarrow \|\rho - \rho_0\|_{\mathsf{op}} \leq \frac{\sigma}{d} \\
&\Rightarrow \|\rho - \frac{1}{d} I_d\|_{\mathsf{op}} \leq \frac{4\sigma}{d} \Rightarrow \rho \in S_{\mathrm{supp}}.
\end{aligned}$$

So, letting X denote the left-hand side of (8), we have

$$\begin{aligned}
X &\geq \int_{B(\rho_0,C\varepsilon)} \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \mathbb{1}\left\{ \|\rho - \rho_0\|_{\mathsf{op}} \leq \frac{C\varepsilon}{d} \right\} \mathrm{dLeb}_U(\rho) \\
&= \int_U \frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \mathbb{1}\{\rho \in N_*(\rho_0)\} \, \mathrm{dLeb}_U(\rho) \\
&= \mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \right] \int_U \mathbb{1}\{\rho \in N_*(\rho_0)\} \, \mathrm{dLeb}_U(\rho).
\end{aligned}$$

Finally, note that

$$\begin{aligned}
&\int_U \mathbb{1}\{\rho \in N_*(\rho_0)\} \, \mathrm{dLeb}_U(\rho) \\
&= \int_{B(\rho_0,C\varepsilon)} \mathbb{1}\left\{ \|\rho - \rho_0\|_{\mathsf{op}} \leq \frac{C\varepsilon}{d} \right\} \mathrm{dLeb}_U(\rho) \\
&= (C\varepsilon)^{(d+2)(d-1)/2} V_2,
\end{aligned}$$

which conludes the proof. □

It remains to control $\mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \right]$ and the volume ratio V_2/V_1. The former measures the information gained from the observations $\boldsymbol{x}$. It is bounded by the following lemma, which formalizes the intuition that each observation improves the log likelihood ratio by at most $O(\varepsilon^2/d)$. This is the step that uses the hypothesis $n \ll d^3/\varepsilon^2$.

Lemma III.9. *If $\rho_0 \in S_{\mathsf{good}}$, then for any sequence of unit vectors $\boldsymbol{x} = (x_1, \ldots, x_n)$,*

$$\mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \right] \geq \exp(-d^2).$$

Proof. For all $\rho \in S_{\mathrm{supp}}$, the eigenvalues of ρ lie within $[0.96/d, 1.04/d]$. Thus, for any unit vector x, $\frac{x^\dagger \rho x}{x^\dagger \rho_0 x} \in [0.96/1.04, 1.04/0.96] \subseteq [0.9, 1.1]$. Using the fact that $\log(1+a) \geq a - \frac{2}{3}a^2$ for $|a| \leq 0.1$, we have

$$\begin{aligned}
\log \frac{x^\dagger \rho x}{x^\dagger \rho_0 x} &\geq \frac{x^\dagger(\rho - \rho_0)x}{x^\dagger \rho_0 x} - \frac{2}{3}\left(\frac{x^\dagger(\rho - \rho_0)x}{x^\dagger \rho_0 x} \right)^2 \\
&\geq \frac{x^\dagger(\rho - \rho_0)x}{x^\dagger \rho_0 x} - d^2 (x^\dagger(\rho - \rho_0)x)^2.
\end{aligned}$$

By symmetry, $\mathbb{E}_{\rho\sim\gamma(\rho_0)}(\rho - \rho_0) = 0$, and by rotational invariance,

$$\begin{aligned}
&d^2 \, \mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[(x^\dagger(\rho - \rho_0)x)^2 \right] = \mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\|\rho - \rho_0\|_F^2 \right] \\
&\leq \mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\|\rho - \rho_0\|_{\mathsf{tr}} \|\rho - \rho_0\|_{\mathsf{op}} \right] \leq \frac{C^2\varepsilon^2}{d}.
\end{aligned}$$

By Jensen's inequality and the above estimates,

$$\begin{aligned}
&\log \mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \right] \geq \sum_{i=1}^n \mathbb{E}_{\rho\sim\gamma(\rho_0)} \log \frac{x_i^\dagger \rho x_i}{x_i^\dagger \rho_0 x_i} \\
&\geq \sum_{i=1}^n \mathbb{E}_{\rho\sim\gamma(\rho_0)} \left[\frac{x_i^\dagger(\rho - \rho_0)x_i}{x_i^\dagger \rho_0 x_i} - d^2 (x_i^\dagger(\rho - \rho_0)x_i)^2 \right] \\
&\geq -\frac{C^2 n \varepsilon^2}{d}.
\end{aligned}$$

Finally, because $n \ll d^3/\varepsilon^2$, this is lower bounded by $-d^2$. □

The volume ratio V_2/V_1 is bounded by the following lemma, whose proof we defer to Section IV. The proof uses tools from random matrix theory.

Lemma III.10. *We have that $V_2/V_1 \geq \exp(-3d^2)$.*

We now put the above claims together to prove Theorem III.6.

Proof of Theorem III.6. Let S_{ρ_0} be the event from Lemma III.7. By the calculation (7) and Lemmas III.7 and III.8, for all $\boldsymbol{x} \in S_{\rho_0}$,

$$\frac{\nu_{\boldsymbol{x}}(B(\rho_0, C\varepsilon))}{\nu_{\boldsymbol{x}}(B(\rho_0, \varepsilon))} \geq \exp(-5d^2) C^{(d+2)(d-1)/2} \, \mathbb{E}_{\rho \sim \gamma(\rho_0)} \left[\frac{\mathrm{d}\mathcal{T}_\rho}{\mathrm{d}\mathcal{T}_{\rho_0}}(\boldsymbol{x}) \right] \cdot \frac{V_2}{V_1}.$$

Lemmas III.9 and III.10 bound the remaining factors, giving

$$\frac{\nu_{\boldsymbol{x}}(B(\rho_0, C\varepsilon))}{\nu_{\boldsymbol{x}}(B(\rho_0, \varepsilon))} \geq \exp(-9d^2) C^{(d+2)(d-1)/2}.$$

Taking $C = e^{20}$ gives $\frac{\nu_{\boldsymbol{x}}(B(\rho_0, C\varepsilon))}{\nu_{\boldsymbol{x}}(B(\rho_0, \varepsilon))} \gg 1$. Since $\nu_{\boldsymbol{x}}(B(\rho_0, C\varepsilon)) \leq 1$, this implies $\nu_{\boldsymbol{x}}(B(\rho_0, \varepsilon)) \ll 1$. □

D. Main Lower Bound

We can now prove our main lower bound for tomography with incoherent measurements, which we state formally below:

Theorem III.11. *There exist absolute constants $\varepsilon_0 > 0$ and $d_0 \in \mathbb{N}$ such that for any $0 < \varepsilon < \varepsilon_0$ and any integer $d \geq d_0$, the following holds. If $n = o(d^3/\varepsilon^2)$, then for any algorithm for state tomography $(\mathcal{T}, \mathcal{A})$ that uses n incoherent, possibly adaptive, measurements, its output $\widehat{\rho}$ upon measuring n copies of ρ satisfies $\|\rho - \widehat{\rho}\|_{\mathsf{tr}} > \varepsilon$ with probability $1 - o(1)$ for some ρ.*

Proof. Let $S \in \sigma(\rho, \boldsymbol{x})$ be the event that $\rho \sim \mu$ lies in S_{good} and $\boldsymbol{x} \sim \mathcal{T}_\rho$ lies in S_ρ. In this proof we will abuse notation and use $\mathcal{A}$ to also denote the internal randomness used by $\mathcal{A}$. It suffices to show $\mathbb{P}_{\mathcal{A}, \rho \sim \mu, \boldsymbol{x} \sim \mathcal{T}_\rho}[\|\mathcal{A}(\boldsymbol{x}) - \rho\|_{\mathsf{tr}} \leq \varepsilon] = o(1)$.

First note that

$$\mathbb{P}_{\mathcal{A}, \rho, \boldsymbol{x}} [\|\mathcal{A}(\boldsymbol{x}) - \rho\|_{\mathsf{tr}} \leq \varepsilon] \tag{9}$$

$$= \mathbb{E}_{\mathcal{A}, \boldsymbol{x}} \, \mathbb{E}_{\rho \sim \nu_{\boldsymbol{x}}} [\mathbb{1}[\|\mathcal{A}(\boldsymbol{x}) - \rho\|_{\mathsf{tr}} \leq \varepsilon]] \tag{10}$$

$$\leq \mathbb{E}_{\mathcal{A}, \boldsymbol{x}} \, \mathbb{E}_{\rho \sim \nu_{\boldsymbol{x}}} [\mathbb{1}[\|\mathcal{A}(\boldsymbol{x}) - \rho\|_{\mathsf{tr}} \leq \varepsilon \text{ and } (\rho, \boldsymbol{x}) \in S]] + o(1) \tag{11}$$

where the second step follows by a union bound and the fact that $\mathbb{P}[(\rho, \boldsymbol{x}) \notin S] = e^{-\Omega(d)} + e^{-\Omega(d^2)} = o(1)$ by Theorem III.6.

For any choice of internal randomness for $\mathcal{A}$ and any transcript $\boldsymbol{x}$, let $\rho_{\boldsymbol{x}}^{\mathcal{A}}$ denote an arbitrary state for which $(\rho_{\boldsymbol{x}}^{\mathcal{A}}, \boldsymbol{x}) \in S$ and $\|\mathcal{A}(\boldsymbol{x}) - \rho_{\boldsymbol{x}}^{\mathcal{A}}\|_{\mathsf{tr}} \leq \varepsilon$, if such a state exists. Denote by $\mathcal{E}$ the event that such a state exists. Then under $\mathcal{E}$, for any state ρ for which $\|\mathcal{A}(\boldsymbol{x}) - \rho\|_{\mathsf{tr}} \leq \varepsilon$, we have $\|\rho_{\boldsymbol{x}}^{\mathcal{A}} - \rho\|_{\mathsf{tr}} \leq 2\varepsilon$. If $\mathcal{E}$ does not occur for some choice of internal randomness for $\mathcal{A}$ and some $\boldsymbol{x}$, note that the corresponding inner expectation in (11) is zero. We can thus upper bound the double expectation in (11) by

$$\begin{aligned} &\mathbb{E}_{\mathcal{A}, \boldsymbol{x} | \mathcal{E}} \, \mathbb{E}_{\rho \sim \nu_{\boldsymbol{x}}} \left[\mathbb{1}[\|\rho_{\boldsymbol{x}}^{\mathcal{A}} - \rho\|_{\mathsf{tr}} \leq 2\varepsilon \text{ and } (\rho, \boldsymbol{x}) \in S] \right] \\ &\leq \mathbb{E}_{\mathcal{A}, \boldsymbol{x} | \mathcal{E}} \, \mathbb{P}_{\rho \sim \nu_{\boldsymbol{x}}} \left[\|\rho_{\boldsymbol{x}}^{\mathcal{A}} - \rho\|_{\mathsf{tr}} \leq 2\varepsilon \right] = o(1), \end{aligned} \tag{12}$$

where in the last step we used the fact that under $\mathcal{E}$ we have $(\rho_{\boldsymbol{x}}^{\mathcal{A}}, \boldsymbol{x}) \in S$, so by Theorem III.6 the posterior measure $\nu_{\boldsymbol{x}}$ places $o(1)$ mass on the trace norm ε-ball around $\rho_{\boldsymbol{x}}^{\mathcal{A}}$. □

IV. Lower Bounding the Volume Ratio: Proof of Lemma III.10

In this section, we will prove Lemma III.10, which lower bounds the volume ratio V_2/V_1.

Let

$$V = \left\{ \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d : \sum_{i=1}^d \lambda_i = 0 \right\}.$$

This is a subspace of $\mathbb{R}^d$ of codimension 1. It inherits the inner product of $\mathbb{R}^d$, which defines a Lebesgue measure Leb_V. Define

$$\Delta = \left\{ \boldsymbol{\lambda} \in V : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d, \sum_{i=1}^d |\lambda_i| \leq 1 \right\},$$

and

$$\Delta' = \left\{ \boldsymbol{\lambda} \in \Delta : \max_{i \in [d]} |\lambda_i| \leq 1/d \right\},$$

and

$$\Gamma = \left\{ \boldsymbol{\lambda} \in V : \left| \lambda_i - \frac{d - 2i + 1}{d^2} \right| \leq \frac{1}{d^4} \right\}.$$

Lemma IV.1. *We have $\Gamma \subseteq \Delta'$.*

Proof. If $\boldsymbol{\lambda} \in \Gamma$, then

$$\lambda_i - \lambda_{i+1} \geq \frac{2}{d^2} - \frac{2}{d^4} > 0$$

and

$$\sum_{i=1}^d |\lambda_i| \leq d \cdot \frac{1}{d} = 1,$$

so $\boldsymbol{\lambda} \in \Delta$. Moreover

$$|\lambda_i| \leq \frac{d-1}{d^2} + \frac{1}{d^4} \leq \frac{1}{d}.$$ □

The volume ratio V_2/V_1 is the probability that if ρ is drawn from the uniform measure on $B(0,1)$ (w.r.t. Leb_{U_0}), then $\|\rho\|_{\mathsf{op}} \leq 1/d$. The main random matrix theory fact we will use is [2, Theorem 2.5.2], which implies that if $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ are the eigenvalues of ρ drawn from this distribution, then $\boldsymbol{\lambda}$ has density (w.r.t. Leb_V) $\frac{1}{Z} f(\boldsymbol{\lambda})$, where Z is a normalizing constant and

$$f(\boldsymbol{\lambda}) = \mathbb{1}\{\boldsymbol{\lambda} \in \Delta\} \prod_{1 \leq i < j \leq d} |\lambda_i - \lambda_j|.$$

For measurable $A \subseteq V$, let $\mathrm{Vol}(A) = \int_A 1 \mathrm{d}\mathsf{Leb}_V$ denote the volume of A. Then,

$$\begin{aligned} \frac{V_2}{V_1} = \frac{\int_{\Delta'} f(\boldsymbol{\lambda}) \, \mathrm{d}\mathsf{Leb}_V(\boldsymbol{\lambda})}{\int_{\Delta} f(\boldsymbol{\lambda}) \, \mathrm{d}\mathsf{Leb}_V(\boldsymbol{\lambda})} &\geq \frac{\int_{\Gamma} f(\boldsymbol{\lambda}) \, \mathrm{d}\mathsf{Leb}_V(\boldsymbol{\lambda})}{\int_{\Delta} f(\boldsymbol{\lambda}) \, \mathrm{d}\mathsf{Leb}_V(\boldsymbol{\lambda})} \\ &\geq \frac{\mathrm{Vol}(\Gamma)}{\mathrm{Vol}(\Delta)} \cdot \frac{\inf_{\boldsymbol{\lambda} \in \Gamma} f(\boldsymbol{\lambda})}{\sup_{\boldsymbol{\lambda} \in \Delta} f(\boldsymbol{\lambda})}. \end{aligned} \tag{13}$$

The following three propositions bound the quantities in the right-hand side of (13).

Proposition IV.2. *For any* $\boldsymbol{\lambda} \in \Gamma$, $f(\boldsymbol{\lambda}) \geq 1/((2e)^{d^2/2} d^{d(d-1)/2})$.

Proof. For each $i \in [d]$,

$$\begin{aligned}
&\prod_{j \in [d] \setminus \{i\}} |\lambda_i - \lambda_j| \\
&\geq \prod_{j=1}^{i-1} \left(\frac{2(i-j)}{d^2} - \frac{2}{d^4} \right) \prod_{j=i+1}^{d} \left(\frac{2(j-i)}{d^2} - \frac{2}{d^4} \right) \\
&\geq \prod_{j=1}^{i-1} \frac{i-j}{d^2} \prod_{j=i+1}^{d} \frac{j-i}{d^2} \\
&\geq \prod_{j=1}^{d-1} \frac{j}{2d^2} \\
&\geq \frac{(d-1)!}{2^d d^{2(d-1)}} = \frac{d!}{2^d d^{2d-1}} \geq \frac{(d/e)^d}{2^d d^{2d-1}} \geq \frac{1}{(2e)^d d^{d-1}}.
\end{aligned}$$

Thus

$$f(\boldsymbol{\lambda}) = \left(\prod_{i=1}^{d} \prod_{j \in [d] \setminus \{i\}} |\lambda_i - \lambda_j| \right)^{1/2} \geq \frac{1}{(2e)^{d^2/2} d^{d(d-1)/2}}. \tag{14}$$

□

Proposition IV.3. *For any* $\boldsymbol{\lambda} \in \Delta$, $f(\boldsymbol{\lambda}) \leq e^{2d^2}/d^{d(d-1)/2}$.

Proof. Let $(\bar{\lambda}_1, \ldots, \bar{\lambda}_d)$ be the permutation of $(\lambda_1, \ldots, \lambda_d)$ with $|\bar{\lambda}_1| \leq \cdots \leq |\bar{\lambda}_d|$. For each $i \in [d]$,

$$\prod_{j<i} |\bar{\lambda}_i - \bar{\lambda}_j| \leq (2|\bar{\lambda}_i|)^{i-1} \leq \left(\frac{e^{2d|\bar{\lambda}_i|}}{d} \right)^{i-1} \leq \frac{e^{2d^2|\bar{\lambda}_i|}}{d^{i-1}},$$

so (since $\sum_{i=1}^{d} |\lambda_i| \leq 1$)

$$f(\boldsymbol{\lambda}) = \prod_{i=1}^{d} \prod_{j<i} |\bar{\lambda}_i - \bar{\lambda}_j| \leq \frac{e^{2d^2 \sum_{i=1}^{d} |\bar{\lambda}_i|}}{d^{d(d-1)/2}} \leq \frac{e^{2d^2}}{d^{d(d-1)/2}}. \square$$

Proposition IV.4. *We have* $\mathrm{Vol}(\Delta) \leq e^{O(d)}$ *and* $\mathrm{Vol}(\Gamma) \geq d^{-O(d)}$.

Proof. The projection of V onto its first $d-1$ coordinates has Jacobian $\Theta(1)$. Because this projection maps Δ injectively to $\mathbb{R}^{d-1}$ and each resulting coordinate is in $[-1, 1]$,

$$\mathrm{Vol}(\Delta) \leq \Theta(1) \cdot 2^{d-1} \leq e^{O(d)},$$

proving the first conclusion. Note that Γ is the set of $\boldsymbol{\lambda}$ satisfying $\lambda_i = \frac{d-2i+1}{d^2} + \frac{1}{d^4}\mu_i$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$ ranges over

$$\Gamma' = \left\{ \boldsymbol{\mu} \in V : \max_{i \in [d]} |\mu_i| \leq 1 \right\}.$$

The set Γ' certainly contains the set of $\boldsymbol{\mu}$ where $|\mu_1|, \ldots, |\mu_{d-1}| \leq 1/d$ and $\mu_d = -\sum_{i=1}^{d-1} \mu_i$. So,

$$\mathrm{Vol}(\Gamma') \geq \Theta(1)(2/d)^{d-1} = d^{-O(d)}.$$

Finally,

$$\mathrm{Vol}(\Gamma) = \mathrm{Vol}(\Gamma') \cdot (d^{-4})^{d-1} = d^{-O(d)},$$

proving the second conclusion. □

Proof of Lemma III.10. By equation (13) and the last three propositions,

$$\frac{V_2}{V_1} \geq \frac{d^{-O(d)}}{e^{O(d)}} \cdot \frac{(2e)^{-d^2/2}}{e^{2d^2}} \geq e^{-3d^2}. \qquad \square$$

V. Basic Learning Results

Recall that for two quantum states ρ, σ, the fidelity between them is $F(\rho, \sigma) = \mathsf{tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2$. This is the quantum analogue of the *Bhattacharyya coefficient*, which for two classical probability distributions p, q over $[d]$, is defined to be $BC(p, q) = \sum_{i=1}^{d} \sqrt{p_i}\sqrt{q_i}$. Note that if ρ and σ commute, then $F(\rho, \sigma) = BC(p, q)^2$, where p, q are the classical distributions given by the eigenvalues of ρ and σ, respectively. The following inequality is well known:

Fact V.1.

$$BC(p, q) \geq 1 - O\left(\chi^2(p\|q)\right) .$$

As an immediate corollary of this, we have the following:

Corollary V.2. *Let* $\rho \in \mathbb{C}^{d \times d}$ *be an arbitrary mixed state, and let* $\tilde{\rho} = (1-\gamma)\rho + \gamma \cdot \frac{1}{d}\mathbb{1}$. *Then* $F(\rho, \tilde{\rho}) \geq 1 - O(\gamma)$.

Proof. Let p and $\tilde{p}$ denote the the distributions given by the eigenvalues of ρ and $\tilde{\rho}$, respectively. Since ρ and $\tilde{\rho}$ commute, it suffices to show that $BC(p, \tilde{p}) \geq 1 - O(\gamma)$. However, we have:

$$\begin{aligned}
\chi^2(\tilde{p}\|p) &= \sum_{i=1}^{d} \frac{(\gamma p_i + \gamma/d)^2}{\tilde{p}_i} \leq 2\sum_{i=1}^{d} \frac{\gamma^2 p_i^2}{\tilde{p}_i} + 2\sum_{i=1}^{d} \frac{\gamma^2}{d^2 \tilde{p}_i} \\
&\leq O(\gamma) ,
\end{aligned}$$

since $\tilde{p}_i \geq \gamma/d$ for all i. □

It is well-known that infidelity is not a metric on mixed states. However, the associated Bures metric, defined as

$$D_B(\rho, \sigma)^2 = 2\left(1 - \sqrt{F(\rho, \sigma)}\right) ,$$

does satisfy the triangle inequality and is hence a valid metric. As an immediate consequence of this, we obtain that infidelity still satisfies a weak version of the triangle inequality:

Corollary V.3. *Let* k *be a positive integer, and let* $\gamma \ll 1/k^2$ *be sufficiently small. Let* $\rho_1, \ldots, \rho_k$ *be a sequence of mixed states satisfying* $1 - F(\rho_i, \rho_{i+1}) \leq \gamma$ *for all* $i = 1, \ldots, k-1$. *Then*

$$1 - F(\rho_1, \rho_k) \leq O(k^2\gamma) .$$

Proof. By repeated application of the Taylor series expansion of the square root function around 1, we know that $D_B(\rho_i, \rho_{i+1}) \leq O(\sqrt{\gamma})$ for all $i = 1, \ldots, k-1$. Therefore by the triangle inequality, we have that $D_B(\rho_1, \rho_k) \leq O(k\sqrt{\gamma})$, from which the claim immediately follows. □

A. Learning a state in spectral norm

Finally, we require the following guarantee for one of the standard estimators for a mixed state based on unentangled measurements. Given n copies of a mixed state ρ, we measure each one with the uniform POVM over the sphere $\{d\,|v\rangle\langle v|\,\mathrm{d}v\}$ where v ranges over the unit sphere, and let $|v_i\rangle$ denote the outcome of the ith measurement. We consider the estimator $H_n(\rho) = H_n(\rho, v_1, \ldots, v_n)$, defined as

$$H_n(\rho) = \frac{1}{n}\sum_{i=1}^{n}\left((d+1)\,|v_i\rangle\langle v_i| - \mathbb{1}\right) . \tag{15}$$

We show the following rate for this estimator. The same rate is claimed in the proof sketch of Theorem 2 in [17], but to our knowledge no full proof of this exists in the literature. We include a full proof for completeness:

Theorem V.4. *There exists a universal constant C so that for all n, we have that*

$$\|H_n(\rho) - \rho\|_{\mathsf{op}} \leq C \cdot \max\left(\frac{d+\log 1/\delta}{n}, \sqrt{\frac{d+\log 1/\delta}{n}}\right) ,$$

with probability $1-\delta$.

The key concentration lemma we require is the following:

Lemma V.5. *Let $|v\rangle$ be the outcome of measuring ρ using the uniform POVM. Then, for any fixed pure state $|u\rangle$, and for all $k \geq 1$, we have*

$$\mathbb{E}\left[(d+1)^k \langle u|v\rangle^{2k}\right] \leq (k+1)^{k+1} .$$

Proof. First consider the case where $\rho = |w\rangle\langle w|$ is a pure state. For any t, let Π_t denote projection onto the t-fold symmetric subspace. Then, we have:

$$\mathbb{E}\left[\langle u|v\rangle^{2k}\right] = d \cdot \int \langle u|v\rangle^{2k}\langle v|w\rangle^2\, dv \tag{16}$$

$$= d \cdot \langle u|^{\otimes k} \otimes \langle w| \left(\int |v\rangle^{\otimes(k+1)} \langle v|^{\otimes(k+1)}\, dv\right) |u\rangle^{\otimes k} \otimes |w\rangle \tag{17}$$

$$= d \cdot \binom{k+d}{k+1}^{-1} \cdot \left(\langle u|^{\otimes k} \otimes \langle w|\right) \Pi_{k+1} \left(|u\rangle^{\otimes k} \otimes |w\rangle\right) \tag{18}$$

$$\leq d \cdot \binom{k+d}{k+1}^{-1} , \tag{19}$$

where the third step follows by the standard Haar integral formulation of Π_t [19]. Therefore, we have that

$$\mathbb{E}\left[(d+1)^k \langle u|v\rangle^{2k}\right] \leq (d+1)^{k+1} \cdot \binom{k+d}{k+1}^{-1} \tag{20}$$

$$\leq (d+1)^{k+1} \cdot \frac{(k+1)^{k+1}}{(d+k)^{k+1}} \tag{21}$$

$$\leq (k+1)^{k+1} , \tag{22}$$

as claimed. The claim for general ρ directly follows by convexity. □

Proof of Theorem V.4. The proof proceeds via the same general strategy as in [17]. Let $\mathcal{N}$ be a $1/3$-net of all pure states in $\mathbb{C}^d$. For any $u \in \mathcal{N}$, Lemma V.5 implies that the random variable

$$\langle u|\left(\rho - H_n(\rho)\right)|u\rangle = \frac{1}{n}\sum_{i=1}^{n}\left((d+1)\langle u, v_i\rangle^2 - 1 - \langle u|\,\rho\,|u\rangle\right)$$

is a sum of n independent $O(1)$-subexponential random variables. Therefore, by standard net arguments, we have that for all $\gamma > 0$,

$$\mathbb{P}[\|\rho - H_n(\rho)\|_{\mathsf{op}} > \gamma] \leq \exp\left(c_1 d - c_2 n \max(\gamma, \gamma^2)\right) ,$$

for some universal constants c_1, c_2, which is equivalent to what we wanted to show. □

Finally, we also require the following generalization of the estimator we considered above:

Definition V.6. *Given a projection matrix $\Pi \in \mathbb{C}^{d\times d}$ onto an r-dimensonal subspace, the* projected estimator *on the subspace Π, denoted $H_n(\rho, \Pi)$ is defined as follows. Let $\mathcal{M}_\Pi = \{r\,|v\rangle\langle v|\,dv\}$ denote the uniform POVM over the subspace spanned by Π, and consider the POVM over $\mathbb{C}^{d\times d}$ defined by $\mathcal{M} = \{I - \Pi\} \cup \mathcal{M}_\Pi$. Given n copies of ρ, let $|v_i\rangle$ denote the outcomes of measuring each copy of ρ independently with $\mathcal{M}$, where we say $v_i = \perp$ if the outcome is $I - \Pi$. Then, we define*

$$H_n(\rho, \Pi) = \frac{1}{n}\sum_{v_i \neq \perp}\left((r+1)\,|v_i\rangle\langle v_i| - \mathbb{1}\right) .$$

For projected estimators we have the following rate, which follows immediately from Theorem V.4 and standard Chernoff bounds:

Lemma V.7. *Let $\rho \in \mathbb{C}^{d\times d}$ and let Π be a projection onto an r-dimensional subspace. Let $\alpha = \mathsf{tr}(\Pi\rho\Pi)$. Then, there exists a universal constant C so that*

$$\|H_n(\rho, \Pi) - \Pi\rho\Pi\|_{\mathsf{op}} \leq C \cdot \max\left(\frac{d+\log 1/\delta}{n}, \sqrt{\frac{\alpha(d+\log 1/\delta)}{n}}\right) .$$

with probability $1-\delta$.

VI. Learning a state in fidelity

In this section, we present our algorithm for tomography in fidelity. Our main theorem is stated below.

Theorem VI.1. *Let $\rho \in \mathbb{C}^{d\times d}$ be an unknown rank-r mixed state. Given $n = O(dr^2\log(1/\delta)/\gamma)$ copies of ρ, there is an algorithm that uses incoherent measurements and with probability $1-\delta$, outputs a state $\widehat{\rho}$ such that $F(\rho, \widehat{\rho}) \geq 1-\gamma$.*

Let $\eta > 0$ be a parameter to be determined later, and let t be an integer satisfying $t \leq \log_2 r/\gamma + 4$. We first describe our algorithm. Divide the samples into t groups of n/t samples each.

We will compute a sequence of orthogonal projections $\Pi_0 = 0, \Pi_1, \ldots, \Pi_t$. To find Π_j given $\Pi_0, \ldots, \Pi_{j-1}$, the algorithm

proceed as follows: it forms $\Gamma_j = \mathbb{1} - \sum_{i=0}^{j-1} \Pi_i$, and using a fresh batch of samples, it computes $\sigma_j = H_{n/t}(\rho, \Gamma_j)$, and it sets Π_j to be the span of the nonzero eigenvectors of σ_j with corresponding eigenvalue at least 2^{-j}. When it finishes, it outputs the state

$$M_n(\rho) \stackrel{\text{def}}{=} \frac{\widehat{\sigma}}{\mathsf{tr}(\widehat{\sigma})} \text{ , where } \widehat{\sigma} \stackrel{\text{def}}{=} \sum_{i=1}^{t} \Pi_i \sigma_i \Pi_i \text{ , } \tag{23}$$

To analyze the algorithm, we make the following definitions. Let $\Pi_{t+1} = \Gamma_{t+1}$, and define the state

$$\rho_{\text{diag}} = \sum_{i=1}^{t+1} \Pi_i \rho \Pi_i \text{ . } \tag{24}$$

For all $i = 1, \ldots, t+1$, let $B_i \in \mathbb{C}^{d \times \text{rank}(\Pi_i)}$ be any matrix with orthonormal columns so that $B_i B_i^\top = \Pi_i$. Similarly, for $i = 1, \ldots, t+1$, let $C_i \in \mathbb{C}^{d \times \text{rank}(\Gamma_i)}$ be any matrix with orthonormal columns so that $C_i C_i^\top = \Gamma_i$.

Notice that by construction $\sum_{i=1}^{t+1} \Pi_i = \mathbb{1}$, so this is indeed a valid mixed state. Furthermore, note that ρ_{diag} is block-diagonal with respect to the matrices $\Pi_1, \ldots, \Pi_{t+1}$, that is, after a suitable rotation which sends each B_i to itself, ρ_{diag} has the form

$$\rho_{\text{diag}} = \begin{pmatrix} \rho_1 & & & \\ & \rho_2 & & \\ & & \ddots & \\ & & & \rho_{t+1} \end{pmatrix} \text{ , }$$

where we let ρ_i denote the projection of $\Pi_i \rho \Pi_i$ onto its nonzero eigenvectors. Note that in this basis, ρ is not block diagonal, and indeed, much of the work in the proof is to bound the contribution of the error because of these off-diagonal blocks. To this end, for all $i = 1, \ldots, t$, let $E_i = B_i^\top \rho C_{i+1}$, so that

$$\rho = \begin{pmatrix} \rho_1 & E_1 & & \\ E_1^\top & \rho_2 & E_2 & \\ & E_2^\top & \ddots & \\ & & & \rho_t \end{pmatrix} \text{ , }$$

i.e., the E_i are the off-diagonal components of ρ.

Next, since each σ_j is computed with fresh samples, by Lemma V.7 and a union bound, we have that

$$\begin{aligned} &\|\sigma_j - \Gamma_j \rho \Gamma_j\|_{\mathsf{op}} \leq \gamma_j \\ &\stackrel{\text{def}}{=} C \cdot \max\left(\frac{d + \log t/\delta}{n}, \sqrt{\frac{\alpha_j (d + \log t/\delta)}{n}} \right) \text{ , } \end{aligned} \tag{25}$$

for all $j = 1, \ldots, t$ simultaneously, with probability at least $1 - \delta$, where $\alpha_j = \mathsf{tr}(\Pi_j \rho \Pi_j)$.

For the remainder of the section, let us take $n \gtrsim \frac{(d + \log t/\delta) r^2}{\gamma}$, and we will assume that (25) holds. We first show a few basic inequalities. We have the following estimate on the RHS of (25).

Lemma VI.2. *Let $n \gtrsim \frac{(d+\log t/\delta) r^2}{\gamma}$. Then, we have that $\gamma_j \lesssim 2^{-(j+t)/2}$.*

Proof. By our choice of n, we have that:

$$\frac{d + \log t/\delta}{n} \lesssim \frac{\gamma}{r^2} \leq 2^{-(j+t)/2} \text{ , }$$

and

$$\begin{aligned} \sqrt{\frac{\alpha_{j-1}(d + \log t/\delta)}{n}} &= \sqrt{\mathsf{tr}(\Gamma_{j-1} \rho \Gamma_{j-1}) \cdot \frac{d + \log t/\delta}{n}} \\ &\lesssim \sqrt{2^{-(j-1)} \cdot \frac{\gamma}{r}} \leq 2^{-(j+t)/2} \text{ , } \end{aligned}$$

where in both cases we use that $t = \log_2 r/\gamma + 4$. □

With this, we can show the following set of useful inequalities:

Lemma VI.3. *For all $j = 1, \ldots, t$, we have that*

(i) $\|\Gamma_j \rho \Gamma_j\|_{\mathsf{op}} \leq 2^{-(j-1)}$,
(ii) $\|\sigma_j\|_{\mathsf{op}} \leq 2^{-(j-2)}$, *and*
(iii) $B_i^\top \rho B_i \succeq 2^{-j-1} \mathbb{1}$, *and*
(iv) $\mathsf{tr}(\Gamma_{t+1} \rho \Gamma_{t+1}) \leq \gamma/2$.

Proof. To prove the first bullet point, we proceed by induction. The case where $j = 1$ is trivial. Now, suppose the claim holds for some $j - 1$. By (25), we have that

$$\|\sigma_{j-1} - \Gamma_{j-1} \rho \Gamma_{j-1}\|_{\mathsf{op}} \leq \gamma_{j-1} \text{ . }$$

Now Γ_j is defined to be the projection onto the eigenvectors of σ_{j-1} with eigenvalue less than 2^{-j}. Therefore, $\|\Gamma_j \rho \Gamma_j\|_{\mathsf{op}} \leq \gamma_{j-1} + 2^{-j}$. The second and third claims then both follow from Lemma VI.2, and by the definition of M_j. Finally, the last claim follows because ρ has at most r nonzero eigenvalues, and therefore so does $\mathsf{tr}(\Gamma_{t+1} \rho \Gamma_{t+1})$; moreover by the above, each one is at most $2^{-t+2} \leq \gamma/(2r)$. □

One simple but important consequence of these inequalities is that the subspaces Π_i are low dimensional:

Corollary VI.4. *For all i, we have* $\text{rank}(\Pi_i) \leq r$.

Proof. Note that ρ has r nonzero eigenvalues by assumption, so by Lemma VI.3, σ_i can only have r eigenvalues which are at least 2^{-i-1}. □

We will show the following two key estimates. Roughly, Lemma VI.5 bounds the contribution to infidelity form the error on the diagonal blocks of ρ and ρ_{diag}. Lemma VI.6 bounds the contribution to infidelity from the off-diagonal blocks in ρ. Putting them together with the (weak) triangle inequality in Corollary V.3 will immediately imply Theorem VI.1.

Lemma VI.5. *Let $\rho, M_n(\rho)$, and ρ_{diag} be as above, and assume that (25) holds. Then, we have that*

$$F(M_n(\rho), \rho_{\text{diag}}) \geq 1 - \gamma t$$

Lemma VI.6. *Let $\rho, M_n(\rho)$, and ρ_{diag} be as above, and assume that (25) holds. Then, we have that*

$$F(\rho, \rho_{\text{diag}}) \geq 1 - \gamma \cdot \min(t, d/r) \text{ . }$$

Proof of Theorem VI.1. Combining Lemma VI.5, Lemma VI.6 and Corollary V.3, we get

$$F(M_n(\rho).\rho) \geq 1 - O(\gamma \min(t, d/r))$$

and since $t \leq \log_2(r/\gamma) + 4$, we can redefine γ appropriately (scaling down by a logarithmic factor) to complete the proof. □

VII. Proof of Lemma VI.5

We first require the following fact, which is a direct consequence of Proposition 3.1 in [23], and direct calculation:

Theorem VII.1. *Let $f(X) : \mathbb{C}^{d\times d}_{\geq 0} \to \mathbb{R}$ be defined by $f(X) = \mathsf{tr}(\sqrt{X})$. Let $A \in \mathbb{C}^{d\times d}$ be a non-degenerate Hermitian matrix. Then, for all symmetric matrices $B \in \mathbb{C}^{d\times d}$, we have that*

$$\nabla^2 f(A)[B, B] \geq -\frac{1}{4}\mathsf{tr}\left(BA^{-3/2}B\right) \tag{26}$$

where $\nabla^2 f(A)$ denotes the Hessian of $f(A)$ (where we treat the entries of A as variables) and $\nabla^2 f(A)[B, B]$ denotes taking the quadratic form of this Hessian at B.

Proof of Lemma VI.5. For conciseness, throughout this proof we will let $M = M_n(\rho)$. Let $\Delta = M - \rho_{\mathrm{diag}}$. By Taylor's theorem, we know that there is some $\beta \in [0, 1]$ so that

$$\begin{aligned}
&F(M_n(\rho), \rho_{\mathrm{diag}}) = \mathsf{tr}\left(\sqrt{M^2 - M^{1/2}\Delta M^{1/2}}\right) \\
&= 1 + \frac{1}{2}\mathsf{tr}\left(M^{-1}M^{1/2}\Delta M^{1/2}\right) + D^2 f \\
&\quad \cdot \left(M^2 - \beta M^{1/2}\Delta M^{1/2}\right)\left[M^{1/2}\Delta M^{1/2}, M^{1/2}\Delta M^{1/2}\right] \\
&\geq 1 - \frac{1}{4}\mathsf{tr}\left((M^{1/2}\Delta M^{-1/4}(M + \beta\Delta)^{-3/2}M^{-1/4}\Delta M^{1/2}\right)
\end{aligned}$$

where in the last line we used Theorem VII.1.

Notice that both M and Δ are block diagonal, and moreover, by construction, they share the same block structure. Thus, if we let $M_i = \Pi_i M \Pi_i$ and $\Delta_i = \Pi_i \Delta \Pi_i$, then we have that

$$\begin{aligned}
&\mathsf{tr}\left((M^{1/2}\Delta M^{-1/4}(M + \beta\Delta)^{-3/2}M^{-1/4}\Delta M^{1/2}\right) \\
&= \sum_{i=1}^{t} \mathsf{tr}\left(M_i^{1/2}\Delta_i M_i^{-1/4}(M_i + \beta\Delta_i)^{-3/2}M_i^{-1/4}\Delta_i M_i^{1/2}\right)
\end{aligned}$$

where in a slight abuse of notation, we let M_i^{-1} and $(M_i + \beta\Delta)^{-1}$ denote the pseudoinverses of the associated matrices. For any i, we have that

$$\begin{aligned}
&\mathsf{tr}\left(M_i^{1/2}\Delta_i M_i^{-1/4}(M_i + \beta\Delta_i)^{-3/2}M_i^{-1/4}\Delta_i M_i^{1/2}\right) \\
&\leq \mathsf{tr}(M_i) \cdot \|\Delta_i M_i^{-1/4}(M_i + \beta\Delta_i)^{-3/2}M_i^{-1/4}\Delta_i\|_{\mathsf{op}} \\
&\leq C \cdot \mathsf{tr}(M_i) \cdot \gamma_i^2 \cdot \|M_i^{-1/4}(M_i + \beta\Delta_i)^{-3/2}M_i^{-1/4}\|_{\mathsf{op}} ,
\end{aligned} \tag{27}$$

where the last step follows because $\|\Delta_i\|_{\mathsf{op}} \leq \gamma_i$ by Lemma V.7 and (25). By construction, we have that all nonzero eigenvalues of M_i are at least 2^{-i}, so all nonzero eigenvalues of $M_i + \beta\Delta_i$ are at least $2^{-(i+1)}$. Putting it all together, we obtain that

$$\begin{aligned}
&\mathsf{tr}\left(M_i^{1/2}\Delta_i M_i^{-1/4}(M_i + \beta\Delta_i)^{-3/2}M_i^{-1/4}\Delta_i M_i^{1/2}\right) \\
&\qquad \lesssim \mathsf{tr}(M_i) \cdot \gamma_i^2 \cdot 2^{2i} \leq r \cdot 2^i\gamma_i^2 \leq \gamma ,
\end{aligned}$$

as claimed. □

VIII. Proof of Lemma VI.6

First, observe that we may disregard the subspace spanned by Γ_{t+1}:

Lemma VIII.1. *Let $\overline{\Gamma} = \mathbb{1} - \Gamma_{t+1}$. We have that $\mathsf{tr}(\overline{\Gamma}\rho\overline{\Gamma}) = \mathsf{tr}(\overline{\Gamma}\rho_{\mathrm{diag}}\overline{\Gamma}) \stackrel{def}{=} c$, and moreover, if we let $\tilde{\rho} = \frac{1}{c}\overline{\Gamma}\rho\overline{\Gamma}$ and $\tilde{\rho}_{\mathrm{diag}} = \frac{1}{c}\overline{\Gamma}\rho_{\mathrm{diag}}\overline{\Gamma}$, then $F(\tilde{\rho}, \rho) \leq 1 - \gamma/2$ and $F(\tilde{\rho}_{\mathrm{diag}}, \rho_{\mathrm{diag}}) \leq 1 - \gamma/2$.*

Proof. The first claim follows since Π_i and $\overline{\Gamma}$ all commute. The fact that $F(\tilde{\rho}, \rho) \leq 1 - \gamma/2$ immediately follows from (iv) in Lemma VI.3, and the last claim follows since the same lemma implies that $\mathsf{tr}(\Gamma_{t+1}\rho_{\mathrm{diag}}\Gamma_{t+1}) \leq \gamma/2$ as well. □

In light of this, for the rest of the proof, we will assume that $\overline{\Gamma} = \mathbb{1}$; in particular, this implies that ρ and ρ_{diag} both have minimum eigenvalue at least 2^{-j-1} by Lemma VI.3. The above lemma implies that this incurs at most an additional additive γ to the overall fidelity calculation. We now establish the following bound on the matrices E_i:

Lemma VIII.2. *Let $n \gtrsim \frac{(d+\log T/\delta)r^2}{\gamma}$, and assume that (25) holds. Then, for all $i = 1, \ldots, t$, we have that $\|E_i\|_{\mathsf{op}} \leq 2\gamma_j$.*

Proof. For any i, we have that

$$\begin{aligned}
\|E_i\|_{\mathsf{op}} &= \|\Gamma_j\left(\rho - \rho_{\mathrm{diag}}\right)\Gamma_j\|_{\mathsf{op}} \\
&\leq \|\Gamma_j\rho\Gamma_j - \sigma_j\|_{\mathsf{op}} + \|\Gamma_j\rho_{\mathrm{diag}}\Gamma_j - \sigma_j\|_{\mathsf{op}} \\
&\leq 2\|\Gamma_j\rho\Gamma_j - \sigma_j\|_{\mathsf{op}} = 2\gamma_j ,
\end{aligned}$$

where the third inequality follows because $\Gamma_j\rho_{\mathrm{diag}}\Gamma_j$ is the projection of $\Gamma_j\rho\Gamma_j$ onto a basis which commutes with σ_j. □

To prove our overall claim we will proceed via a hybrid argument. For all $j = 0, \ldots, t$, let

$$\rho^{(j)} = \sum_{i=1}^{j} \Pi_i\rho\Pi_i + \Gamma_{j+1}\rho\Gamma_{j+1} .$$

Note that by design, we have that $\rho^{(0)} = \rho$ and $\rho^{(t)} = \rho_{\mathrm{diag}}$. For these matrices, we show:

Lemma VIII.3. *For all $j = 1, \ldots, t$, we have that*

$$F(\rho^{(j-1)}, \rho^{(j)}) \geq 1 - 2^{-t-1}\min(rt, d) .$$

To prove Lemma VIII.3, we rely on the following fact that bounds the contribution from off-diagonal perturbations to infidelity between two states.

Lemma VIII.4. *Let $0 < c_2, c_1 < 1$ satisfy $c_2 \leq c_1/10$, and let $M \in \mathbb{C}^{d_1\times d_1}$ and $N \in \mathbb{C}^{d_2\times d_2}$ be PSD matrices satisfying*

$c_1\mathbb{1} \preceq M \preceq 4c_1\mathbb{1}$ *and* $c_2\mathbb{1} \preceq N \preceq 2c_1\mathbb{1}$. *Let* $E \in \mathbb{C}^{d_1\times d_2}$ *satisfy* $\|E\|_{\mathsf{op}} \leq \eta$, *and define the matrices*

$$A = \begin{pmatrix} M & E \\ E^\top & N \end{pmatrix} \text{ , and } A_{\mathrm{diag}} = \begin{pmatrix} M & 0 \\ 0 & N \end{pmatrix} . \tag{28}$$

Suppose further that $A \succeq 0$, *and that* $\eta \leq \sqrt{c_1 c_2}$. *Then*

$$\mathsf{tr}\left(\sqrt{A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2}}\right) \geq \mathsf{tr}(A) - c_2(d_1+d_2) . \tag{29}$$

Proof. For any δ, let $M_\delta = M - \delta\mathbb{1}$ and similarly let $N_\delta = N - \delta\mathbb{1}$. we will explicitly construct a matrix B which will be a PSD lower bound for the matrix $A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2}$. Our guess will have the form

$$B = \begin{pmatrix} M_{c_2} & X \\ X^\top & N_{c_2} \end{pmatrix} , \tag{30}$$

for some matrix X we define shortly. Note that if we can show that $A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2} \succeq B^2$, we are done, since by operator monotonicity of the matrix square root, we have that

$$\begin{aligned} \mathsf{tr}\left(\sqrt{A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2}}\right) &\geq \mathsf{tr}(B) \\ &\geq \mathsf{tr}(A) + \mathsf{tr}(M) - c_2(d_1+d_2) \\ &= \mathsf{tr}(A) - c_2(d_1+d_2) , \end{aligned}$$

as claimed.

It remains to demonstrate how to construct such an X. We choose the following matrix:

$$X \stackrel{\text{def}}{=} \sum_{i=1}^{\infty} M_{(c_2-c_1)}^{-i} \left(M^{1/2} E N^{1/2}\right) N_{(c_2+c_1)}^{i-1} . \tag{31}$$

We first note that this sum is convergent, since $\|M_{c_2-c_1}^{-i}\| \leq \frac{1}{(1.9c_1)^i}$ and $\|N_{c_2+c_1}^i\|_{\mathsf{op}} \leq (1.1c_1)^i$. Next, we note that

$$M_{c_2}X + XN_{c_2} = M_{(c_2-c_1)}X + XN_{(c_2+c_1)} = M^{1/2}EN^{1/2} .$$

Therefore, we have that

$$\begin{aligned} B^2 &= \begin{pmatrix} M_{c_2}^2 + XX^\top & M_{c_2}X + XN_{c_2} \\ (M_{c_2}X + XN_{c_2})^\top & X^\top X + N_{c_2}^2 \end{pmatrix} \\ &= \begin{pmatrix} M_{c_2}^2 + XX^\top & M^{1/2}EN^{1/2} \\ (M^{1/2}EN^{1/2})^\top & X^\top X + N_{c_2}^2 \end{pmatrix} , \end{aligned} \tag{32}$$

At the same time, we have that

$$A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2} = \begin{pmatrix} M^2 & M^{1/2}EN \\ (M^{1/2}EN^{1/2})^\top & N^2 \end{pmatrix} , \tag{33}$$

so in particular the off-diagonal block of B^2 exactly matches that of $A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2}$. Therefore,

$$A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2} - B^2 \tag{34}$$

$$= \begin{pmatrix} M^2 - M_{c_2}^2 - XX^\top & \\ & N^2 - N_{c_2}^2 - X^\top X \end{pmatrix} \tag{35}$$

$$= \begin{pmatrix} 2c_2 M - c_2^2 I - XX^\top & \\ & 2c_2 N - c_2^2 I - X^\top X \end{pmatrix} \tag{36}$$

Therefore for our candidate to be a valid PSD lower bound, it suffices to show that $c_2^2 I + XX^\top \preceq 2c_2 M$ and similarly $c_2^2 I + X^\top X \preceq 2c_2 N$. Note that for both of these inequalities to be satisfied, it suffices to show that $X^\top X \preceq c_2 N$.

Define the matrix

$$Q = \sum_{i=1}^{\infty} M_{(c_2-c_1)}^{-i} \left(M^{1/2}E\right) N_{(c_2+c_1)}^{i-1} . \tag{37}$$

Since N and N_δ commute for all δ, we have that

$$X^\top X = N^{1/2} Q^\top Q N^{1/2} .$$

Additionally, we have that

$$\|Q\|_{\mathsf{op}} = \left\| \sum_{i=1}^{\infty} M_{(c_2-c_1)}^{-i} \left(M^{1/2}E\right) N_{(c_2+c_1)}^{i-1} \right\|_{\mathsf{op}} \tag{38}$$

$$\leq \sum_{i=1}^{\infty} \left\| M_{(c_2-c_1)}^{-i} \left(M^{1/2}E\right) N_{(c_2+c_1)}^{i-1} \right\|_{\mathsf{op}} \tag{39}$$

$$\leq \frac{1.9}{1.1c_1} \cdot \|M^{1/2}E\|_{\mathsf{op}} \sum_{i=1}^{\infty} \left(\frac{1.1}{1.9}\right)^i \tag{40}$$

$$\lesssim \frac{\eta}{\sqrt{c_1}} \leq \sqrt{c_2} , \tag{41}$$

by assumption. Therefore, we have that $X^\top X \preceq c_2 N$, so B is indeed a valid PSD lower bound on $A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2}$. □

Proof of Lemma VIII.3. Fix some $j \in \{1,\ldots,t\}$. Let $B = C_{j+1}^\top \rho C_{j+1}$, and define the matrices

$$A \stackrel{\text{def}}{=} \begin{pmatrix} \rho_j & E_j \\ E_j^\top & B \end{pmatrix} , \text{ and } A_{\mathrm{diag}} = \begin{pmatrix} \rho_j & \\ & B \end{pmatrix} .$$

Then

$$F(\rho^{(j-1)}, \rho^{(j)}) = F\left(\begin{pmatrix} \tau & \\ & A \end{pmatrix}, \begin{pmatrix} \tau & \\ & A_{\mathrm{diag}} \end{pmatrix}\right) ,$$

for some shared, unnormalized state τ. Recall that by Lemma VI.2, we have that $\rho_j \succeq 2^{-j-1}\mathbb{1}$, and that $2^{-t-1}\mathbb{1} \preceq B \preceq 2^{-j+1}\mathbb{1}$. Additionally, we have that $\|E_j\|_{\mathsf{op}} \leq 2\gamma_j$. Therefore the matrices A and A_{diag} satisfy the conditions of Lemma VIII.4 with parameters $c_1 = 2^{-j-1}$, $c_2 = 2^{-t-1}$, and $\eta = 2\gamma_j$. Note that Lemma VI.2 immediately implies that $\eta \leq \sqrt{c_1 c_2}$, as necessary. Therefore we have that

$$F\left(\begin{pmatrix} \tau & \\ & A \end{pmatrix}, \begin{pmatrix} \tau & \\ & A_{\mathrm{diag}} \end{pmatrix}\right) \tag{42}$$

$$= \mathsf{tr}(\tau) + \mathsf{tr}\left(\sqrt{A_{\mathrm{diag}}^{1/2} A A_{\mathrm{diag}}^{1/2}}\right) \tag{43}$$

$$\geq \mathsf{tr}(\tau) + \mathsf{tr}(A) - 2^{-t-1}\left(\sum_{i=1}^{t} \mathrm{rank}(\Pi_i)\right) \tag{44}$$

$$\geq 1 - 2^{-t-1}\min(rt, d) , \tag{45}$$

as claimed. □

We can now finish the proof of Lemma VI.6.

Proof of Lemma VI.6. The desired statement follows by combining Lemma VIII.1, Lemma VIII.3 and Corollary V.3 and redefining γ appropriately (scaling down by a logarithmic factor). □

IX. Conclusion

In this work we obtained the optimal lower bound of $\Omega(d^3/\varepsilon^2)$ for state tomography in trace distance using adaptive incoherent measurements. Our proof technique is based on setting up a suitable Bayesian problem and arguing that the posterior distribution over the underlying state, conditioned on the measurement outcomes, places negligible mass on the ground truth if $o(d^3/\varepsilon^2)$ measurements are made. This technique is a marked departure from existing approaches to proving lower bounds for adaptive incoherent measurements, which exclusively applied to the *testing* setting. We leave as an open question whether one can refine this lower bound to $\Omega(dr^2/\varepsilon^2)$ in the case where the unknown state is rank-r.

Additionally, for the problem of state tomography in infidelity, we gave an *adaptive* algorithm that achieves the optimal rate of d^3/γ, up to logarithmic factors. By a lower bound of [18], this rate is superior to that of any nonadaptive algorithm. Our algorithm uses logarithmically many rounds of adaptivity, and we conjecture that this much adaptivity is necessary to achieve the optimal rate for incoherent measurements.

a) Acknowledgments.: The authors would like to thank Steve Flammia and Ryan O'Donnell for coordinating their submission of [14] with ours. SC and JL would like to thank Jordan Cotler and Hsin-Yuan Huang for illuminating discussions about tomography with incoherent measurements.

References

[1] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature communications*, 13(1):1–9, 2022.

[2] Greg Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2005.

[3] Costin Bădescu, Ryan O'Donnell, and John Wright. Quantum state certification. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 503–514, 2019.

[4] Emilio Bagan, MA Ballester, Richard D Gill, Alex Monras, and Ramon Munoz-Tapia. Optimal full estimation of qubit mixed states. *Physical Review A*, 73(3):032301, 2006.

[5] Konrad Banaszek, Marcus Cramer, and David Gross. Focus on quantum tomography. *New Journal of Physics*, 15(12):125020, 2013.

[6] Sebastien Bubeck, Sitan Chen, and Jerry Li. Entanglement is necessary for optimal quantum property testing. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 692–703. IEEE, 2020.

[7] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. A hierarchy for replica quantum advantage. *arXiv preprint arXiv:2111.05874*, 2021.

[8] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585. IEEE, 2022.

[9] Sitan Chen, Brice Huang, Jerry Li, and Allen Liu. Tight bounds for quantum state certification with incoherent measurements. *arXiv preprint arXiv:2204.07155*, 2022.

[10] Sitan Chen, Jerry Li, and Ryan O'Donnell. Toward instance-optimal state certification with incoherent measurements. *arXiv preprint arXiv:2102.13098*, 2021.

[11] Omar Fawzi, Nicolas Flammarion, Aurélien Garivier, and Aadil Oufkir. On Adaptivity in Quantum Testing. working paper or preprint, May 2023.

[12] Omar Fawzi, Nicolas Flammarion, Aurélien Garivier, and Aadil Oufkir. Quantum channel certification with incoherent strategies. *arXiv preprint arXiv:2303.01188*, 2023.

[13] Christopher Ferrie and Robin Blume-Kohout. Minimax quantum tomography: the ultimate bounds on accuracy. *arXiv preprint arXiv:1503.03100*, 2015.

[14] Steve Flammia and Ryan O'Donnell. Quantum chi-squared tomography and mutual information testing. *preprint*, 2023.

[15] Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012.

[16] Christopher A Fuchs and Jeroen Van De Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999.

[17] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020.

[18] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Theory*, 63(9):5628–5641, 2017.

[19] Aram W Harrow. The church of the symmetric subspace. *arXiv preprint arXiv:1308.6595*, 2013.

[20] Masahito Hayashi and Keiji Matsumoto. Asymptotic performance of optimal state estimation in qubit system. *Journal of Mathematical Physics*, 49(10):102101, 2008.

[21] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R McClean. Quantum advantage in learning from experiments. *arXiv preprint arXiv:2112.00778*, 2021.

[22] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021.

[23] Anatoli Juditsky and Arkadii S Nemirovski. Large deviations of vector-valued martingales in 2-smooth normed spaces. *arXiv preprint arXiv:0809.0813*, 2008.

[24] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

[25] Angus Lowe. Learning quantum states without entangled measurements. Master's thesis, University of Waterloo, 2021.

[26] Dylan H Mahler, Lee A Rozema, Ardavan Darabi, Christopher Ferrie, Robin Blume-Kohout, and AM Steinberg. Adaptive quantum state tomography improves accuracy quadratically. *Physical review letters*, 111(18):183601, 2013.

[27] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

[28] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 529–538, 2015.

[29] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 899–912, 2016.

[30] Ryan O'Donnell and John Wright. Efficient quantum tomography II. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 962–974, 2017.

[31] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *arXiv preprint arXiv:2008.02269*, 2020.

[32] Joran van Apeldoorn, Arjan Cornelissen, András Gilyén, and Giacomo Nannicini. Quantum tomography using state-preparation unitaries. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1265–1318. SIAM, 2023.